العنوان:	التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات
المؤلف الرئيسي:	عبدالله، سمية عبداالله محمد الأمين
مؤلفين آخرين:	أحمد، شاذلى صديق محمد(مشرف)
التاريخ الميلادي:	2017
موقع:	الخرطوم
الصفحات:	1 - 82
رقم MD:	916054
نوع المحتوى:	رسائل جامعية
اللغة:	Arabic
الدرجة العلمية:	رسالة ماجستير
الجامعة:	جامعة النيلين
الكلية:	كلية علوم الحاسوب وتقانة المعلومات
الدولة:	السودان
قواعد المعلومات:	Dissertations
مواضيع:	التنبؤ بالمبيعات، شركة سويتش للتجارة والهندسة، قسم المولدات، تقنية أشجار القرار، التنقيب عن البيانات، قواعد البيانات
رابط:	http://search.mandumah.com/Record/916054

© 2024 المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو المنظومة.

للإستشهاد بهذا البحث قم بنسخ البيانات التالية حسب إسلوب الإستشهاد المطلوب:

إسلوب APA

عبداالله، سمية عبداالله محمد الأمين، و أحمد، شاذلى صديق محمد. (2017).التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات(رسالة ماجستير غير منشورة). جامعة النيلين، الخرطوم. مسترجع من http://search.mandumah.com/Record/916054

إسلوب MLA

عبداالله، سمية عبداالله محمد الأمين، و شاذلى صديق محمد أحمد. "التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات" رسالة ماجستير. جامعة النيلين، الخرطوم، 2017. مسترجع من http://search.mandumah.com/Record/916054

© 2024 المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو المنظومة.

الباب الثالث

(تحليل الخوارزميات)

3-1 التحليل

يعدُّ تحليل البيانات وتفسيرها خطوةً موصِّلة إلى النتائج، بعد إتمام تجهيز البيانات وتصنيفها إلى مرحلة تحليلها وتفسيرها واختبار فرضيَّاتها لاستخلاص النتائج منها وتقدير إمكانيَّة تعميمها؛ أي أنّهَ لكي يتم التوصل إلى ذلك نحتاج إلى تحليل بيانات، وقد كان تحليل المعلومات والبيانات حتى وقت قريب يقتصر على التحليل الفلسفيّ والمنطقيّ والمقارنة البسيطة، ولكنَّ الاتِّجاه في الوقت المعاصر هو الاعتماد على الطرق الإحصائيَّة والأساليب الكميَّة؛ فهي تساعد على تحليل بيانات الدراسة ووصفها وصفاً أكثر دقَّة، وتساعد على حساب الدقَّة النسبيَّة للقياسات المستخدمة،وتعدُّ مرحلة التحليل من أهمِ المراحل وأخطرها، وعليها تتوقَّف التفسيرات والنتائج؛ [17]

3-2تعريف المشكلة (تحليل المشكلة)

وجود كم هائل من البيانات لدى قسم التسويق خاص بمبيعات السنين السابقة وعدم وجود تقنية للاستفادة من تلك البيانات لمعرفة سلوك الزبائن الشرائية .

2. ايجاد تقنية للتنبؤ بمتطلبات السوق في المستقبل حسب سعات المنتج

3-3.مرحلة جمع البياناتData Collection

اولا كانت الـ data ware house (data marts) هي عبارة عن بيانات عمليات البيع التي تمت من كل العملاء لشركة سويتش والذين قاموا بشراء مولدات منذ بداية توكيل مولدات SDMO الفرنسية عام 1990وحتى 2017 تم جمع البيانات في شكل جداول حسابية (Microsoft Excel File)وهي عبارة عن 11104 سجل (record) تم رصدها لدى إدارة التسويق سابقا وقد اخذنا منها لاغراض هذه الدراسة 1100 سجل (في الفترة من 2015 -9/2017) بعد ازلة التكرار المطبعي والضجيج (القيم الغير معلومة اوالتي لاتنتمي لما هو معروف من اسماء او سعات او عناوين) واستبعاد التاريخ القديم والموديلات التي لم تعد تتعامل معها الشركة وتنفيذ كل عمليات data clean اصبحث 1000 سجل.

3-4 وصف البيانات Dataset Description :

تم تحديد خواص ومواصفات كل متغير مثل نوع المتغير واسمه.

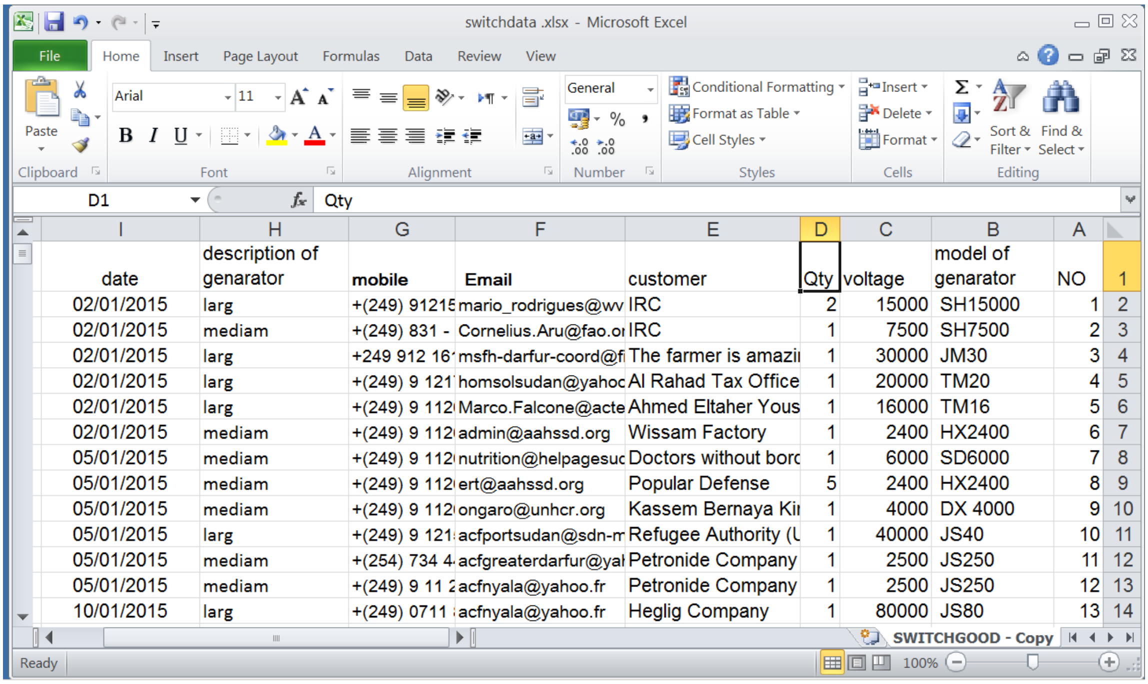
جدول رقم (5) يوضح خواص ومواصفات كل متغير

Attribute	type	desc
NO	Numeric	**الرقم**
model of genarator	Nominal	**موديل المولد**
voltage	Numeric	**سعة المولد**
Qty	Numeric	**الكمية**
customer	Nominal	**العميل**
Email	Nominal	بريد العميل
mobile	Nominal	موبايل العميل
description of genarator	Nominal	وصف حجم المولد
date	Nominal	تاريخ الشراء

3-5البرنامج المستخدم:

البرنامج المستخدم في هذه الدراسة والذي عن طريقه سوف يتم تصميم شجرة القرار هو برنامج WEKA.

جدول رقم (6) يوضح مجموعة البيانات لمبيعات المولدات ماركة **SDMO** شركة سويتش للتجارة والهندسة في الفترة من 2015 وحتى الربع الثاني من 2017



NO	model of genarator	voltage	Qty	customer	Email	mobile	description of genarator	date
1	SH15000	15000	2	IRC	mario_rodrigues@wv	+(249) 91215	larg	02/01/2015
2	SH7500	7500	1	IRC	Cornelius.Aru@fao.oı	+(249) 831 -	mediam	02/01/2015
3	JM30	30000	1	The farmer is amazi	msfh-darfur-coord@fi	+249 912 16	larg	02/01/2015
4	TM20	20000	1	Al Rahad Tax Office	homsolsudan@yahoc	+(249) 9 121	larg	02/01/2015
5	TM16	16000	1	Ahmed Eltaher Yous	Marco.Falcone@acte	+(249) 9 112	larg	02/01/2015
6	HX2400	2400	1	Wissam Factory	admin@aahssd.org	+(249) 9 112	mediam	02/01/2015
7	SD6000	6000	1	Doctors without borc	nutrition@helpagesuc	+(249) 9 112	mediam	05/01/2015
8	HX2400	2400	5	Popular Defense	ert@aahssd.org	+(249) 9 112	mediam	05/01/2015
9	DX 4000	4000	1	Kassem Bernaya Kii	ongaro@unhcr.org	+(249) 9 112	mediam	05/01/2015
10	JS40	40000	1	Refugee Authority (L	acfportsudan@sdn-m	+(249) 9 121	larg	05/01/2015
11	JS250	2500	1	Petronide Company	acfgreaterdarfur@yal	+(254) 734 4	mediam	05/01/2015
12	JS250	2500	1	Petronide Company	acfnyala@yahoo.fr	+(249) 9 11 2	mediam	05/01/2015
13	JS80	80000	1	Heglig Company	acfnyala@yahoo.fr	+(249) 0711	larg	10/01/2015

هذا الجدول يحتوي علي مجموعة البيانات التي رصدها من قسم التسويق بشركة سويتش للتجارة والهندسة _وقد الغي منها عمود الاسعار نسبة لخصوصية المعلومة لدى الشركة_ وهي تمثل الجزء الذي تم عليه التصنيف عدد (1000) سجل وعدد الصفات (9) وهي الرقم, العميل , بريد العميل , رقم موبايل العميل , موديل المولد , سعة المولد , تصنيف حجم المولد, تاريخ الشراء, الكمية.

3-6 تجهيز البيانات Dataset processing :

اخذت البيانات من قسم التسويق ولكن تجهيز البيانات أخذ وقت طويل من زمن عملية التنقيب , وذلك لانها كانت في ملفات متفرقة وكانت هنالك بيانات غير مكتملة ومشوشة تم استبعادها ويوجد تكرار احيانا واخطاء مطبعية مما قد يخفي أنماط مفيدة .

تم تجهيز تلك البيانات بالشكل الملائم لعملية التنقيب , ثم وضعت البيانات على هيئة CSV وذلك لان برنامج Weka يتعامل مع إمتدادات معينة مثل *.CSV & *.ARFF

3 -7 تنقية البيانات Dataset Filter

وهي مرحلة المعالجة الأولية للبيانات.

كانت هنالك مشاكل مختلفة للبيانات وتم حلها:

حل مشكلة البيانات الناقصة Missing Value(وجود حقل فارغ Null ليس به قيم) بالمتوسط للقيم الرقمية واخذ الانسب والاقرب قيمة للقيم الحرفية, معظم عمليات التنقيب تفترض عدم وجود قيم ناقصة.

كما تم مسح البيانات التي بها شوائب(غير مناسبة).

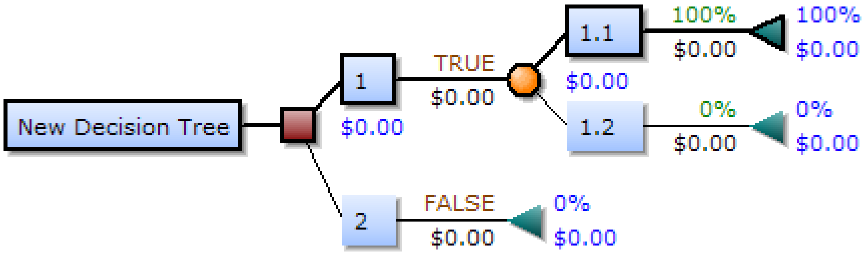
3-8 شرح خطوات الخوارزمية شجرة القرار (J48) :

شجرة اتخاذ القرارات تعتبر إحدى الآليات البسيطة المستخدمة في الذكاءالاصطناعي من أجل الوصول لقرار محدد بناء على المعطيات, تركيب الشجرة يتكون من مجموعة من العناصر تسمى العقد، وهذه العقد ترتبط فيما بينها بواسطة روابط تسمى الحواف. بشكل عام فإننا نسمى أي تركيب يتألف من عقد وحواف بالرسم البياني graph فإذا ترتبت هذه العقد والحواف على شكل جذر وفروع فإنها تصبح شجرة.

شجرة القرار هي هيكلية شبيهة بالمخطط الانسيابي حيث كل عقدة داخلية تمثل "اختبار" للخاصية (مثال هل الوجه الظاهر للعملة المعدنية هو طرة أم نقش)، كل فرع يمثل مخرجات الاختبار وكل عقدة نهائية تمثل القرار المتخذ بعد احتساب جميع الخصائص. المسار من الجذر إلى العقدة النهائية يمثل قواعد التصنيف.

تتكون شجرة القرار من ثلاث أنواع من العقد:

1. عقد القرار – يتم تمثيلها عادة بمربعات.
2. عقد الفرص (الاحتمال) – يتم تمثيلها بدوائر.
3. العقد النهائية – يتم تمثيلها بمربعات.



شكل رقم (15)المخطط الانسيابي لخوارزمية J48(decision Tree)

9-3 استخدام خوارزمية شجرة القرار في التحليل:

تم استخدام مصنف شجرة القرار التي تنفذ الخوارزمية C4.5 وهي تستخدم أسلوب فرق تسد من أعلى إلى أسفل (top-down recursive divide-and-conquer manner) لتوليد أشجار القرار التي تبدأ بالتدريب على مجموعة بيانات التدريب والفئات (class labels) المرتبطة بها , يتم تقسيم مجموعة التدريب بشكل متكرر إلى مجموعات فرعية أصغر ريثما يجري بناء الشجرة.

للخوارزمية ثلاث معلمات

1. **قسم البيانات** : مجموعة كاملة من بيانات التدريب والفئات (class labels) المرتبطة بها
2. **قائمة الصفة** : قائمة من الصفات التي تصف السجلات
3. **مقاييس اختيار الصفة** : هي إرشادات لاختيار معيار التقسيم الأفضل الذي يقسم بيانات معينة (switchdata) , مقاييس اختيار السمة تعرف أيضا باسم قواعد التقسيم لأنها تحدد كيفية تقسيم السجلات في عقدة معينة.

مقاييس اختيار الصفة توفر تصنيف لكل صفة تصف الصف المعين , الصفة التي لديها أفضل مجموع علامات يتم اختيارها كصفة لتقسيم الصف المعين توجد ثلاثة معايير عامة لاختيار الصفة المناسبة للتقسيم الأفضل.

الخوارزمية

إنشاء شجرة القرارات , تولد شجرة القرار من مجموعة سجلات التعلم الموجودة في الملف(switchdata.csv)

المدخلات

قسم البيانات الموجودة في الملف(switchdata.csv) وهي مجموعة من سجلات التعلم أدخل

قائمة الحقول وهي مجموعة من الصفات المرشحة

طريقة اختبار الصفة كانت **Percentage Split** حيث تم تقسيم مجموعة التدريب الموجودة في الملف(switchdata.csv) إلى قسمين مجموعة يتم تدريبة عليها بنسبة 66% (تم الاختيار عشوائيا

بنسبة 1:2من المفترض ان تعطينا اي نسب مختارة اجابات متقاربة) ومجموعة يتم اختباره عليها بنسبة 34%(النسبة المتبقية من 100%)

الإخراج : شجرة قرار

الطريقة :

1. إنشاء عقدة N.
2. إذا كانت كل السجلات في D من نفس الفئة C ، إذن
3. أرجع N كعقدة ورقية تحمل قيمة الفئة C
4. إذا كانت قائمة الصفة فارغة , إذن
5. أرجع N كعقدة ورقية تحمل قيمة الفئة العظمى في D
6. طبق طريقة اختيار الصفة (D,قائمة الصفة) لاختيار أفضل معيار تقسيم
7. ضع الاختبار في العقدة N
8. إذا كانت صفة التقسيم قيمة متقطعة والتقسيمات متعددة الاتجاهات مسموح بها // لا تقتصر على الأشجار الثنائية , إذن
9. قائمة الصفة ←——— قائمة الصفة – صفة التقسيم // إزالة سمة التقسيم
10. لكل نتائج معيار تقسيم J // قسم السجلات وأبني الأشجار الفرعية لكل قسم
11. أجعل Dj تكون مجموعة سجلات البيانات في D التي تحقق J // قسم
12. إذا كانت Dj فارغة إذن
13. إرفاق ورقة تحمل فئة الأغلبية في D إلى العقدة N
14. أو إرفاق العقدة التي أرجعها توليد شجرة القرار (Dj , قائمة الصفة) إلى العقدة N

نهاية لكل

15. أرجع N